MULTIMEDIA UNIVERSITY

STUDENT IDENTIFICATION NO

MULTIMEDIA UNIVERSITY

FINAL EXAMINATION

TRIMESTER 1, 2018/2019

DSP5018 – STATISTICS AND PROBABILITY

(*for diploma students only*)

16 OCTOBER 2018
2.30 p.m – 4.30 a.m
(2 Hours)

INSTRUCTIONS TO STUDENT

1. This question paper consists of **3** pages excluding cover page and appendix.
2. Attempt **ALL FOUR (4)** questions. The marks distribution for each question is given.
3. Write your answers in the answer booklet provided.
4. Key formulae are given in the Appendix.

QUESTION 1

a. The following table shows the results of a survey done on 20 MMU students to see their preference of smartphone brand.

Xiaomi	Samsung	Samsung	Xiaomi	Motorola
Samsung	Xiaomi	Motorola	iPhone	iPhone
Samsung	Xiaomi	Motorola	Motorola	Motorola
Samsung	iPhone	Motorola	Xiaomi	Motorola

i. Construct frequency and relative frequency table. (3 marks)

ii. What is the mode of these data? (1 mark)

iii. Identify the type of variable involved in the survey. (1 mark)

b. Mr Harris recorded the marks of his student midterm exam for the Business Mathematics subject as follows:

46	29	35	82	37	38	45	44	62	31

i. Based on table above, list down the 5 number summary of the midterm exam marks. (4 marks)

ii. Calculate the interquartile range, and identify the lower and upper limit of the distribution. (3 marks)

iii. Identify any outlier(s) and draw a boxplot to represent the above data. (4 marks)

iv. If the mark of the outlier(s) is NOT included, what is the mean marks for the midterm exam? (3 marks)

c. Table below shows the record of transaction amount per day at Burger Rack outlet for 40 days. Calculate the standard deviation of transaction amount per day for the 40 days. (6 marks)

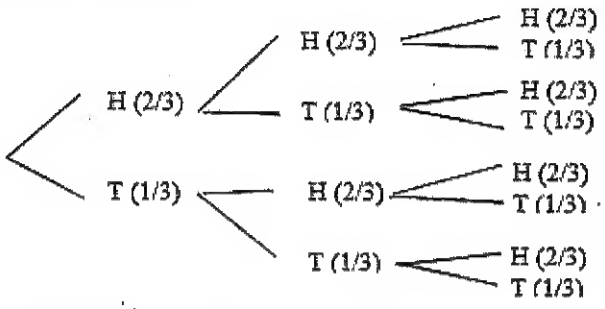
Transaction Amount (RM)	**Number of days**
$0 \leqslant 20$	2
$20 \leqslant 40$	7
$40 \leqslant 60$	11
$60 \leqslant 80$	10
$80 \leqslant 100$	8
$100 \leqslant 120$	2

(Total: 25 marks)

Continued...

QUESTION 2

a. A special coin is designed to have 3 sides where 2 of the sides are head (H) and the other side is tail (T). The coin is tossed 3 times, and the following tree diagram illustrate all the possible outcome.



i. Calculate the probability to get all head in 3 toss. (2 marks)

ii. What is the probability to get 1 tail? (4 marks)

iii. What is the probability of getting at least 1 head? (3 marks)

b. There were 100 students registered on the first day of registration for March 2018 intake in MMU Melaka. The new students registered for two subjects as follow:

	Accounting	**English**
Male	35	y
Female	15	y

i. Calculate the value of y. (2 marks)

ii. Find the probability that a randomly selected student is a male or registered in English subject. (4 marks)

iii. Calculate the probability that a randomly selected student is a female given that she registered in English subject. (3 marks)

c. An overtime claim by employees of a private company are normally distributed with a mean of RM100 and a standard deviation of RM4 per month. Find the probability that an employee will claim:

i. more than RM104 per month. (4 marks)

ii. more than RM87 per month. (3 marks)

(Total: 25 marks)

Continued...

QUESTION 3

a. A machine dispenses 250mg candies into its packaging. A sample of 9 packages are taken and the average weight is 250.0056mg with a standard deviation of 0.0246. Assuming a normal population, determine a 99% confidence interval for the mean weight of the packages. (4 marks)

b. The Kopionomy Café manager would like to know the average amount spent by a customer per visit. A sample of 49 customers over a 10 days period was randomly selected and the average amount spent was RM18.90 with standard deviation of RM4.50. Using 1% level of significance, can he conclude that the average amount spent by a customer is less than RM20 per visit? (10 marks)

c. A study among 609 students in MMU found that 329 of them exercise at least 2 times a week. Can it be concluded at 5% significance level, that the percentage of MMU students exercise at least 2 times a week is not 50%? (11 marks)

(Total: 25 marks)

QUESTION 4

a. A resort operator would like to study the usage of electricity in one of his dorm facilities. Maximum number of occupants at one time is 40 pax. Below are the electricity records for 9 previous groups of occupants.

Number of occupants, x	22	26	27	29	34	33	29	40	30
Electricity usage (kW), y	490	540	555	606	661	675	577	804	738

i. Given $\sum y^2 = 3{,}623{,}496, \sum x^2 = 8{,}316, \sum xy = 173{,}161$. Compute the value of SS_{xx}, SS_{yy} and SS_{xy}. (4 marks)

ii. Calculate the correlation coefficient, r, and interpret the answer. (3 marks)

iii. Compute the regression equation, $\hat{y} = \beta_0 + \beta_1 x$ (5 marks)

iv. If customer come with 35 pax capacity, estimate how much electrical usage (in kW) will they use. (2 marks)

v. Find all the sum of squares, *SST*, *SSR* and *SSE*. (3 marks)

vi. At 5% significance level, do the data provide sufficient evidence to conclude that the number of occupants and amount of electrical usage are positively linearly correlated? (8 marks)

(Total: 25 marks)

End of page.

APPENDIX – KEY FORMULA

	Ungrouped Data	Grouped Data
Mean	$\bar{x} = \frac{\sum x_i}{n}$ where n : sample size	$\bar{x} = \frac{\sum m_i f_i}{\sum f_i}$ where m : class midpoint, f : class frequency
Variance	$s^2 = \frac{1}{n-1}\left[\sum x_i^2 - \frac{(\sum x_i)^2}{n}\right]$ where n : sample size	$s^2 = \frac{1}{(\sum f_i)-1}\left[\sum m_i^2 f_i - \frac{(\sum m_i f_i)^2}{\sum f_i}\right]$ where m : class midpoint, f : class frequency

NORMAL AND STANDARD NORMAL PROBABILITY DISTRIBUTION

- z-value (observed value) for an x value : $Z = \frac{x-\mu}{\sigma}$

ESTIMATION

The $(1-\alpha)100\%$ confidence interval for population mean	
$\bar{x} \pm \left(z_{\frac{\alpha}{2}}\right)(\sigma_{\bar{x}})$ where $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$	$\bar{x} \pm \left(t_{\frac{\alpha}{2},n-1}\right)(s_{\bar{x}})$ where
or	$s_{\bar{x}} = \frac{s}{\sqrt{n}}$
$\bar{x} \pm \left(z_{\frac{\alpha}{2}}\right)(s_{\bar{x}})$ where $s_{\bar{x}} = \frac{s}{\sqrt{n}}$	degree of freedom, $df = n - 1$

The $(1-\alpha)100\%$ confidence interval for population proportion
$\hat{p} \pm \left(z_{\frac{\alpha}{2}}\right)(s_{\hat{p}})$ where $\hat{p} = \frac{x}{n}$, $s_{\hat{p}} = \sqrt{\frac{\hat{p}\hat{q}}{n}}$, $\hat{q} = 1 - \hat{p}$

Continued...

HYPOTHESIS TESTING

Population	Mean	Proportion
1	$Z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$	$Z = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$ where $\hat{p} = \dfrac{x}{n}$
2	$Z = \dfrac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$	$Z = \dfrac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\hat{p}_p(1-\hat{p}_p)}\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$ where $\hat{p}_p = \dfrac{x_1 + x_2}{n_1 + n_2}$

LINEAR REGRESSION ANALYSIS

- **Least square regression equation,** $\hat{y} = \beta_0 + \beta_1 x$ where $\beta_1 = \dfrac{S_{xy}}{S_{xx}}$ and $\beta_0 = \bar{y} - \beta_1 \bar{x}$
- **Correlation coefficient,** $r = \dfrac{S_{xy}}{\sqrt{S_{xx} S_{yy}}}$ where

$S_{xx} = \sum x^2 - \dfrac{(\sum x)^2}{n}$	$S_{yy} = \sum y^2 - \dfrac{(\sum y)^2}{n}$	$S_{xy} = \sum xy - \dfrac{(\sum x)(\sum y)}{n}$

- **Sum of Square**

$SST = SSR + SSE$	$SST = S_{yy}$	$SSR = \dfrac{(S_{xy})^2}{S_{xx}}$

- **Test of Significance for Regression Slope**

Standard Error, $s_e = \sqrt{\dfrac{SSE}{n-2}}$	Regression t-Test $t = \dfrac{\beta_1}{s_e / \sqrt{Sxx}}$	Correlation t-Test $t = r \cdot \sqrt{\dfrac{n-2}{1-r^2}}$

End of page.